Claims

1. A method for detecting moving video objects in a compressed digital bitstream which represents a sequence of fields or frames of video information for one or more previously captured scenes of video, comprising the steps of:
   a. analyzing said compressed bitstream to locate scene cuts therein, thereby determining at least one sequence of fields or frames of video information which represents a single video scene;
   b. estimating one or more operating parameters for a camera which initially captured said video scene by analyzing a portion of said compressed bitstream which corresponds to said video scene; and
   c. detecting one or more moving video objects represented in said compressed bitstream by applying global motion compensation with said estimated operating parameters .

2. The method of claim 1, further comprising the step of extracting visual features of said one or more detected moving video objects from said compressed bitstream.

3. The method of claim 1, wherein said compressed bitstream comprises a bitstream compressed in accordance with the MPEG video standard.

4. The method of claim 1, wherein said analyzing step further comprises the steps of:
   a. parsing said compressed bitstream into blocks of video information and associated motion vector information for each field or frame of video information which comprises the determined sequence of fields or frames of video information representative of said single scene;

b. performing inverse motion compensation on each of said parsed blocks of video information to derive discrete cosign transform coefficients for each of said parsed blocks of video information;

c. counting said motion vector information associated with each of said parsed blocks of video information; and

d. Determining from said counted motion vector information and said discrete cosign transform coefficient information whether one of said scene cuts has occurred.

5. The method of claim 1, wherein said analyzing step comprises parsing said compressed bitstream into blocks of video information and associated motion vector information for each field or frame of video information which comprises the determined sequence of fields or frames of video information representative of said single scene, and wherein said estimating step comprises the step of estimating any zoom and any pan of said camera by determining a multi-parameter transform model applied to said parsed motion vector information.

6. The method of claim 5, wherein said estimating step comprises the steps of:

a. computing each parameter for a multi-parameter affine transform which represents a transformation from a current frame of video information to a previous frame of video; and

b. computing said multi-parameter affine transform to thereby determine global motion information representative of said zoom and pan of said camera.

7. The method of claim 6, wherein said detecting step comprises computing local object motion for said one or more moving video objects based on said global motion information and on one or more of said motion vectors which correspond to said one or more moving video objects.

8. The method of claim 7, further comprising the steps of :

a. determining whether said local object motion is greater than a predetermined threshold;

b. applying morphological operations to said determined local object motion values to eliminate any erroneously sensed moving objects; and

c. determining border points of said detected moving objects to thereby locate a bounding box for said detected moving object.

9. An apparatus for detecting moving video objects in a compressed digital bitstream which represents a sequence of fields or frames of video information for one or more previously captured scenes of video, comprising:

a. means for analyzing said compressed bitstream to locate scene cuts therein and to determine at least one sequence of fields or frames of video information which represents a single video scene;

b. means, coupled to said analyzing means, for estimating one or more operating parameters for a camera which initially viewed said video scene by analyzing a portion of said compressed bitstream which corresponds to said video scene; and

c. means, coupled to said estimating means, for detecting one or more moving video objects represented in said compressed bitstream by applying global motion compensation with said estimated operating parameters.

10. The apparatus of claim 9, further comprising means, coupled to said detecting means, for extracting visual features of said one or more detected moving video objects from said compressed bitstream.

11. The apparatus of claim 9, wherein said compressed bitstream comprises a bitstream compressed in accordance with the MPEG video standard, and wherein said analyzing means further comprises:

a. parsing means for receiving and parsing said compressed bitstream into blocks of video information and associated motion vector information for each field or frame of video information which comprises the determined sequence of fields or frames of video information representative of said single scene;

b. means, coupled to said parsing means, for performing inverse motion compensation on each of said parsed blocks of video information to derive discrete cosign transform coefficients for each of said parsed blocks of video information;

c. counting means, coupled to said inverse motion compensation means, for counting said motion vector information associated with each of said parsed blocks of vide information; and

d. determining means, coupled to said counting mens, for determining from said counted motion vector information and said discrete cosign transform coefficient information whether one of said scene cuts has occurred.

12. The apparatus of claim 9, wherein said analyzing means further comprises means for parsing said compressed bitstream into blocks of video information and associated motion vector information for each field or frame of video information which comprises the determined sequence of fields or frames of video information representative of said single scene, and wherein said estimating means further comprises means for estimating any zoom and any pan of said camera by determining a multi-parameter transform model applied to said parsed motion vector information.

13. The apparatus of claim 12, wherein said estimating means further comprises:

a. means for computing each parameter for a multi-parameter affine transform which represents a transformation from a current frame of video information to a previous frame of video; and

b. means, coupled to said transform parameter computing means, for computing said multi-parameter affine transform to thereby determine global motion information representative of said zoom and pan of said camera.

14. The apparatus of claim 12, wherein said detecting means further comprises means for computing local object motion for said one or more moving video objects based on said global motion information and on one or more of said motion vectors which correspond to said one or more moving video objects.

15. The apparatus of claim 14, further comprising:

a. comparison means, coupled to said local object motion computing means, for determining whether said local object motion is greater than a predetermined threshold;

b. morphological operation means, coupled to said comparison means, for determined local object motion values to eliminate any erroneously sensed moving objects; and

c. border point determination means, coupled to said morphological operation means, for determining border points of said detected moving objects to thereby locate a bounding box for said detected moving object.

16. A method for dissolving an incoming scene of video information which comprises a sequence of fields or frame of compressed video information and an outgoing scene of video information which comprises a sequence of fields or frame of compressed video information, comprising the steps of:

a. applying DCT domain motion inverse compensation to obtain DCT coefficients for all blocks of video information which make up a last frame of said outgoing video scene;

b. applying DCT domain inverse motion compensation to obtain the DCT coefficients for all blocks of video information which make up the first frame of said incoming video scene; and

c. creating a first frame in a dissolve region from said DCT coefficients of said last outgoing frame and said first incoming frame.

17. The method of claim 16, further comprising the step of choosing an initial value for a weighing function prior to step (c).

18. The method of claim 17, further comprising the steps of

d. incrementing said weighting function value; and

e. creating a second frame in said dissolve region from said DCT coefficients of said last outgoing frame and said first incoming frame using said incremented weighing function value.

19. A method for masking a region of a compressed frame of digital video information, comprising the steps of :

a. determining whether said frame to be masked is intra-coded, predictive-coded or bi-directionally predictive-coded;

b. if said frame is intra-coded:

i. extracting DCT coefficients for all blocks within said frame;

ii. examining $block_n$ to determine where in said frame said block is located;

iii. setting said DCT coefficients for said block to zero if said block is outside said mask region;

iv. applying a DCT cropping algorithm to said DCT coefficients if said block is on the boundary of said mask region; and

v. repeating steps (b)(ii) - (b)(iv) for each block in said frame;

c. If said frame is predictive-coded or bi-directionally predictive-coded:

i. examining motion vectors associated with $block_n$ to determine whether they point to blocks outside or on said mask region;

ii. reencoding said block if a motion vector points to blocks outside or on said mask region; and

iii. repeating steps (c)(i) - (c)(ii) for all blocks in said frame.

20. A method for generating a reduced speed sequence of frames of video information from a sequence of frames of compressed video information, comprising the steps of:

a. selecting a frame of compressed video information to be repeated;

b. determining whether said frame to be repeated is intra-coded, predictive-coded or bi-directionally predictive-coded;

c. converting said frame into an intra-coded frame if said frame is a predictive-coded or bi-directionally predictive-coded frame;

d. creating duplicate predictive-coded frames; and

e. arranging said determined frame and said duplicate predictive-coded frames into a sequence of compressed frames of video information.

21. The method of claim 20, wherein said reduced speed sequence of frames generates a frozen frame effect.

22. The method of claim 20, further comprising the step of converting said frame into an intra-coded frame if said frame is a predictive-coded or bi-directionally predictive-coded frame.

23. A system for editing compressed video information over a distributed network, comprising:

a. a client computer;

b. a network link, coupled to said client computer, for permitting said client computer to search for and locate compressed video information on said distributed network; and

c. means for editing a compressed bitstream of video information over said distributed network.

24. The system of claim 21, wherein said editing means includes means for dissolving an incoming scene of video information which includes a sequence of fields or frame of compressed video information and an outgoing scene of video information which includes a sequence of fields or frame of compressed video information, said dissolving means comprising:

a. outgoing motion compensation means for applying DCT domain motion compensation to a last frame of said outgoing video scene to obtain DCT coefficients for all blocks of video information which make up said last frame of said outgoing video scene;

b. incoming motion compensation means, for applying DCT domain motion compensation to a first frame of said incoming video scene to obtain the DCT coefficients for all blocks of video information which make up said first frame of said incoming video scene; and

c. dissolve region creating means, coupled to said incoming motion compensation means and to said outgoing motion compensation means, for creating a first frame in a dissolve region from said DCT coefficients of said last outgoing frame and said first incoming frame.

25. The system of claim 21, wherein said editing means includes means for masking a region of a compressed frame of digital video information, said masking means comprising:

a. means for determining whether said frame to be masked is intra-coded, predictive-coded or bi-directionally predictive-coded;

b. means for processing intra-coded frames including:

i. means for receiving DCT coefficients and for extracting DCT coefficients for all blocks within said frame;

ii. means, coupled to said receiving means, for examining blocks of compressed video information in said frame to determine where in said frame said block is located;

iii. means, coupled to said examining means, for setting said DCT coefficients for said block to zero if said block is outside said mask region; and

iv. means, coupled to said setting means, for applying a DCT translation algorithm to said DCT coefficients if said block is on the boundary of said mask region; and

c. means for processing predictive-coded and bi-directionally predictive-coded frames, including:

i. means for receiving and examining motion vectors associated with blocks of video information to determine whether they point to blocks outside or on said mask region; and

ii. means, coupled to said receiving means, for reencoding said block if a motion vector points to blocks outside or on said mask region.

26. The system of claim 21, wherein said editing means includes means generating a reduced speed sequence of frames of video information from a sequence of frames of compressed video information, comprising:

a. selection means for selecting a frame of compressed video information to be repeated;

b. computational means, coupled to said selection means, for determining whether said frame to be repeated is intra-coded, predictive-coded or bi-directionally predictive-coded;

c. frame generating means, coupled to said computational means and to said converting means, for creating duplicate predictive-coded frames; and

d. frame arranging means, coupled to said frame generating means, for arranging said determined frame and said duplicate predictive-coded frames into a sequence of compressed frames of video information.

27. A method for converting a full resolution compressed domain bitstream into a reduced resolution compressed bitstream, comprising the steps of:

a. examining a frame of compressed video information from said full resolution bitstream;

b. determining whether said examined frame is intra-coded, predictive-coded or bi-directionally predictive-coded;

c. extracting DCT DC coefficients for said determined frame if said determined frame is intra-coded

d. applying DCT domain inverse motion compensation to said frame to extract DCT DC coefficients if said frame is predictive-coded or bi-directionally predictive-coded; and

e. converting said extracted DCT DC coefficients into DCT DC coefficients for a reduced size intra-coded frame of video.